THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 58, No. 2, February 1979
Printed in U.S.A.

Properties of Kruithof's Projection Method

By R. S. KRUPP

(Manuscript received July 24, 1978)

In 1937, J. Kruithof introduced a scheme for projecting from measured point-to-point teletraffic data to some future values, based upon estimates of total originating and terminating traffic only. This study seeks to give a unified picture of Kruithof's projection method and its generalizations, with some practical details and recommendations for implementation. The main text deals with existence and convergence testing, treatment of ill-conditioned or slowly convergent cases, and various extensions of the basic method. An appendix includes proofs of existence, uniqueness, convergence, and continuity.

I. INTRODUCTION

J. Kruithof's method[1] for projecting from measured point-to-point teletraffic data q_{ij} to some future values p_{ij} is based upon estimates of total originating and terminating traffic only. While the original publication (in Flemish) did not receive as much attention as it may deserve, the idea was good enough that it has been independently reinvented numerous times in the intervening years. Related techniques have turned up in economics, statistics, biophysics, pattern recognition, and vehicular traffic studies, for instance. Such repetition largely seems due to a scientific "Babel" effect: workers in different technical disciplines can no longer read each other's work and recognize the same problem in a new context.

While Kruithof showed that his method had certain properties which are clearly desirable in a projection scheme, he did not investigate the underlying mathematical problems. Subsequent workers, such as Bear,[2] Kullback,[3] Sinkhorn,[4,5] Theil,[6] and particularly Csiszar,[7] have thrown much light on these matters. This paper seeks to give a unified picture of Kruithof's method and its many generalizations, with some practical details and recommendations for implementation. Much of the more intricate mathematics is relegated to an appendix, including proofs of existence, uniqueness, convergence, and continuity. These are cited as needed in the main text, which contains information on existence and convergence testing, treatment of ill-conditioned and slowly convergent cases, and various extensions of the basic method. The final comments propound a rationale for Kruithof projection in the context of Bell System planning.

II. KRUITHOF'S BASIC METHOD

In 1937, J. Kruithof[1] proposed a technique for predicting the point-to-point traffic p_{ij} in a given year from a number of originating points $i = 1, 2 \cdots M$ to a number of terminating points $j = 1, 2 \cdots N$. The units could be calls, trunks, or erlangs, for instance, as long as it makes sense to add up various entries in the matrix $\mathbf{p} = [p_{ij}]$. It is assumed that the corresponding traffic matrix $\mathbf{q} = [q_{ij}]$ is known for some other year, while total traffic b_i at each point i and d_j at each j have already been estimated by some external means to yield:

$$b_i = \sum_j p_{ij} \tag{1}$$

$$d_j = \sum_i p_{ij}. \tag{2}$$

Then Kruithof's formula for projecting $\mathbf{p}$ from $\mathbf{q}$ is:

$$p_{ij} = q_{ij} E_i F_j. \tag{3}$$

The "growth factors" E_i and F_j for the originating and terminating points are implicitly defined, and must be computed by solving (1)–(3) simultaneously.

Kruithof recommended that (1)–(3) be solved by starting with the estimate $\mathbf{p} = \mathbf{q}$, then alternately normalizing the rows of $\mathbf{p}$ to satisfy (1) and the columns to satisfy (2), until it stops changing. In practice, this scheme suffers from a tendency to accumulate roundoff error. A mathematically equivalent procedure with better numerical properties would be to substitute (3) into (1) and (2), solving for E_i and F_j to obtain:

$$E_i = b_i / \sum_j q_{ij} F_j \tag{4}$$

$$F_j = d_j / \sum_i q_{ij} E_i. \tag{5}$$

Starting from an arbitrary estimate, such as $F_j = 1$ for all j, (4) and (5) may be evaluated alternately until $\mathbf{p}$ converges.

Various questions arise naturally in connection with this projection method:

(*i*) Under what conditions do (1)–(3) possess a solution $\mathbf{p}^*$?
(*ii*) Can there be more than one solution for $\mathbf{p}^*$?
(*iii*) How does $\mathbf{p}^*$ vary with the estimates of total traffic?
(*iv*) Does the iteration converge, and if so, to what?

(*v*) When should we stop iterating a given case?
(*vi*) Are there valid generalizations of this scheme?

Many of these points are treated at considerable length in the appendix, which discusses a generalized problem: find a probability distribution **P** which satisfies arbitrary linear constraints, such as (1)–(2), and is related to a given distribution **Q** by a product formula, such as (3). To make the connection in the present case, our first step is to divide **p** by the sum $\hat{p}$ of all its elements, reducing it to a joint probability $\mathbf{P} \equiv \mathbf{p}/\hat{p}$ that a call, trunk, or other increment of traffic is from i to j. Now (1) and (2) yield relations:

$$B_i \equiv b_i/\hat{p} = \sum_j P_{ij} \tag{1'}$$

$$D_j \equiv d_j/\hat{p} = \sum_i P_{ij}, \tag{2'}$$

with B_i and D_j the marginal distributions of traffic on i and j. Similarly, **q** is divided by the sum $\hat{q}$ of its elements to get the joint distribution $\mathbf{Q} \equiv \mathbf{q}/\hat{q}$. Let the events $\{e\}$ in the appendix be the set of pairs $e = (i, j)$ and the constraints $\{c\}$ corresponding to (29) be (1′)–(2′) for all points i and j. Then taking $E_i = V_i\hat{p}/\hat{q}$ and $F_j = W_j$ puts the projection formula (3) in exactly the product form (32):

$$P_{ij} = Q_{ij}E_iF_j\hat{q}/\hat{p} = Q_{ij}V_iW_j, \tag{3'}$$

so that we have a case of the general Kruithof problem defined in the appendix.

Now the existence conditions from the appendix show that there is a solution $\mathbf{p}^*$ of the form (3) if and only if (1) and (2) have some solution p_{ij} that vanishes for each q_{ij} that vanishes and is positive whenever q_{ij} is positive. The uniqueness results show there is at most one solution $\mathbf{p}^*$. The solution is continuous in all b_i and d_j whenever it exists. The iteration on (4) and (5) can be recognized as an example of a relaxation procedure, as discussed and analyzed in the appendix. If (1)–(2) possess a solution p_{ij} that is zero for each q_{ij} that is zero, the iteration will converge to some **p** with these same properties. The limit may not be of the form (3) though, since p_{ij} can also vanish for some $q_{ij} > 0$. When a solution $\mathbf{p}^*$ to (1)–(3) exists, however, the iteration converges to it uniquely. This explains the resistance of (4)–(5) to roundoff effects, since such perturbations die out in the process of converging. In the special case that q_{ij} is symmetric and $b_i = d_i$ for all i, uniqueness shows that p_{ij} is also symmetric, since $\mathbf{p}^*$ and its transpose both satisfy (1)–(3). We should note that only $\mathbf{p}^*$ is unique, not the factors E_i and F_j in (3). For instance, replacing them by E_i/a and aF_j for all i, j, and any $a > 0$ will produce the same $\mathbf{p}^*$. The extent of this nonuniqueness is characterized later.

Kruithof pointed out two desirable properties possessed by the projection scheme. The first, called "reversibility," says that after projecting traffic from **q** at time 1 to **p** at time 2, we can turn around and project backward to time 1, recovering **q** exactly. The second property, "divisibility," says projecting from **q** to **p** and then from **p** to **r** at time 3 yields just the same result as projecting directly from **q** to **r**. Thus the projection scheme is not noisy, in the sense of irretrievably losing information along the way about the initial traffic. Rather, **p** depends only on **q** and the row and column sums, not on the path followed over time. Kruithof also described a third desirable property, "separability," which did not hold for his basic method. The idea is to be able to merge or split a collection of points i or j, without affecting the projected traffic for any other points. By careful generalization of Kruithof's method, a property similar to this can be introduced.

III. NETWORK FLOW CONSIDERATIONS

An important aspect of Kruithof's method may be visualized by means of a flow on the simple directed graph in Fig. 1. The nodes $i = 1, 2 \cdots M$ and $j = 1, 2 \cdots N$ represent originating and terminating points. The edge joining node i to node j carries traffic p_{ij}. Interpreting (1)–(2) as conservation laws, the remaining edges to source s and sink t carry total traffic quantities b_i and d_j, while the net flow from s to t is the sum $\hat{p}$ of all the flows p_{ij}. When an element q_{ij} vanishes in **q**, the corresponding edge from i to j is deleted in the network, so that flow p_{ij} is automatically zero. For the flow **p** to have the form (3), all remaining edges must have nonzero flow p_{ij} on them. Conversely, if a flow p_{ij} can be constructed that satisfies (1)–(2) and does not vanish on any edge of the network, then it fulfills the existence conditions, so that (1)–(3) have a solution $\mathbf{p}^*$.

The labeling method of Ford and Fulkerson[8] immediately springs to mind as a means of constructing a flow p_{ij} to satisfy (1)–(2). This is a simple, efficient, easily programmed algorithm that maximizes the net flow from s to t. We just assign maximum capacities of b_i for edges from s to i, infinity for edges from i to j, and d_j for edges from j to t. If the maximum flow obtained is less than:

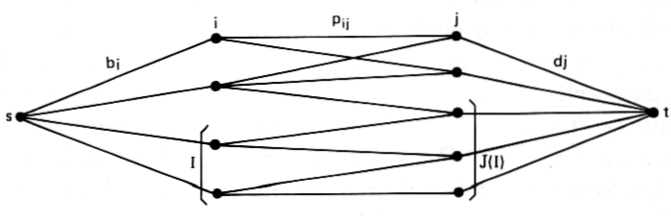

Fig. 1—Network model.

$$\hat{p} = \sum_i b_i = \sum_j d_j, \tag{6}$$

then no such solution of (1)–(2) exists. The only trouble is that the flow obtained may not be positive on all edges, as required to establish that (1)–(3) have a solution. This can be cured by assigning a sufficiently small lower bound $p_{ij} \geqslant h > 0$ to flow on the edges from i to j. An initial flow of h on each such edge will then be feasible.

In practice, the total traffic quantities b_i and d_j are integers or can be scaled up and rounded off to an approximate integer problem without loss of credibility. The labeling method may then be modified to carry h as an infinitesimal quantity, while preserving pure integer arithmetic for efficiency and to avoid roundoff error. Indeed, since one wants to draw a yes-or-no conclusion about existence of a solution, roundoff introduces an unwelcome uncertainty. In the modified scheme, all capacities and flow variables are carried as pairs (m,n) of integers, representing the expression $m + nh$. In the steps of the basic labeling algorithm,[8] one adds and subtracts various capacities and flows, attaching labels to nodes according to whether or not the result exceeds zero. In the modification, the obvious rules apply for adding and subtracting quantities $m + nh$; it remains, however, to specify an interpretation for inequalities involving such quantities. Two classes of inequalities must be defined:

(*i*) $m + nh > 0(1)$ means $m \geqslant 1$.
(*ii*) $m + nh > 0(h)$ means $m \geqslant 1$ or $m = 0$ but $n \geqslant 1$.

In designating h an infinitesimal, we are really promising to choose it as small as necessary so that $|nh| < 1$ for all n that arise.

The solution proceeds in two phases. First the standard labeling method is used, but with all inequality tests to be taken as $> 0(1)$. Thus, in each iteration, an augmenting path of capacity $> 0(1)$ is found. The net flow increases by at least one unit (plus or minus some nh), so that this phase terminates after a finite number of iterations. At this point, the net flow must be $\hat{p} - \hat{n}h$ for some $\hat{n} \geqslant 0$, or else no flow solution of (1)–(2) exists; in effect, we have the maximum flow for the special case $h = 0$. The second phase repeats the standard labeling method, but with all inequality tests to be taken as $> 0(h)$. Now each augmenting path has capacity $> 0(h)$, increasing net flow by at least h units and, again, the iterations terminate in a finite number of steps. If the maximum flow reaches $\hat{p}$, then a solution $\mathbf{p}$ of (1)–(2) has been constructed that fulfills the existence conditions; otherwise, no such solution exists. The two-phase algorithm was programmed, for the network associated with an arbitrary $M \times N$ matrix $\mathbf{q}$, in about eighty lines of Fortran. In tests, it was able to settle the question of existence of solutions to specific Kruithof problems with gratifying rapidity. The same technique can be used to maximize flow on any network under any mixture of $>$ and $\geqslant$ capacities.

Still more can be gleaned from the network model above. Observe that in (6) the sum of the b_i must equal the sum of the d_j in order to conserve flow. This requirement is just a simple example of a class of necessary conditions arising from (1)–(3). More generally, let I be any subset of originating nodes i and $J(I)$ be the subset of terminating nodes j with edges to nodes in I. That is, node j is in $J(I)$ if q_{ij} is positive for some node i in I. Then the flow b_i to node i in I can only pass to nodes in $J(I)$, so that it is included in the flows d_j out of these nodes. Thus the total flow $Y(I)$ from nodes in $J(I)$ cannot be less than the total flow $X(I)$ to all the nodes in I:

$$X(I) \equiv \sum_{I} b_i \leqslant Y(I) \equiv \sum_{J(I)} d_j \tag{7}$$

for each proper subset I (that is, I not empty and not containing all i). Conversely, these necessary conditions guarantee that every cut has a capacity of at least $\hat{p}$, so that (6)–(7) are also sufficient conditions for existence of a flow p_{ij} that satisfies (1)–(2).

Accounting for eq. (3), the requirement of positive flow on each edge allows conditions (7) to be strengthened. Indeed, flow $Y(I)$ from $J(I)$ includes the flow $X(I)$ into I plus any additional flow to $J(I)$ from nodes i that are not in I. If any edge joins a node outside I to $J(I)$, then its flow is positive and (7) becomes a strict inequality:

$$X(I) < Y(I). \tag{8}$$

If equality holds in (7), the rest of the network has no connection to I and $J(I)$, except through s and t. Such a disconnected situation represents two or more independent Kruithof problems, which ought to be treated separately from the outset. Indeed, the rows and columns of q_{ij} can then be renumbered so that it is partitioned into two or more uncoupled blocks. To simplify the statement of later results, we assume that the problem does not decompose in this way, so that the network is connected and (8) holds for every proper subset I of the nodes i. In complete analogy to the case of (6)–(7), necessary conditions (6)–(8) are also sufficient for existence of a positive flow satisfying (1)–(2), and hence for existence of $\mathbf{p}^*$. To see this, we first reduce all network flows and capacities by the initial feasible flow h used in the modified labeling method. For sufficiently small $h > 0$, conditions (7)–(8) now apply to the reduced b_i and d_j as well. But these again show enough capacity on every cut that a flow $p_{ij} - h \geqslant 0$ exists satisfying (1)–(2).

Conditions (6)–(8) can also be used in a direct proof that $\mathbf{p}^*$ exists, independently of results in the appendix. The idea is to seek a stationary point of the quotient *Num*/*Den* of two multinomials in the growth factors $F_j \geqslant 0$. The numerator is positive on the interior and vanishes at boundaries (that is, where some $F_j = 0$):

$$Num \equiv \prod_j F_j^{d_j}, \tag{9}$$

while the denominator is also positive on the interior:

$$Den \equiv \prod_i \left[\sum_j q_{ij} F_j \right]^{b_i}. \tag{10}$$

Setting a derivative of Num/Den with respect to F_j to zero yields the same result as substituting (4) into (5). Thus, if the values F_j make Num/Den stationary, they will also yield a solution $\mathbf{p}^*$ of (1)–(3). Note that Num and Den are both homogeneous of order $\hat{p}$ from (6), so that their quotient is positive and constant along interior rays (that is, along aF_j for all $a > 0$). Since the rays form a compact set and Num/Den is positive and continuously differentiable on the interior, it is enough to show that the quotient goes to zero on the boundary to deduce that it achieves a (stationary) interior maximum. Now suppose that Den vanishes at some boundary point $\mathbf{F}$ and let I be the set of nodes i for which the factor $\sum q_{ij} F_j$ in (10) is zero. Then F_j must vanish if q_{ij} is positive for some i in I, and hence for every j in $J(I)$. If Z measures distance from an interior point $\mathbf{F}'$ to boundary point $\mathbf{F}$, then the numerator will vanish as $\mathbf{F}'$ approaches $\mathbf{F}$ at least as fast as $Z^{Y(I)}$, while the denominator goes like $Z^{X(I)}$, from (7). Thus Num/Den approaches zero at each boundary point $\mathbf{F}$, from (8), completing the proof. The enterprising reader may find it instructive to ferret out the connection between the preceding proof of existence and the more general proofs in the appendix.

One immediate consequence is that the Kruithof problem always has a solution $\mathbf{p}^*$ if $\mathbf{q}$ is strictly positive and (6) holds. Indeed, each node i has an edge to every node j, so that $Y(I) = \hat{p}$ for every proper I and (8) follows. To verify this case more directly, one can construct the positive flow $p_{ij} = b_i d_j / \hat{p}$, which satisfies (1)–(2). Another useful example is a square matrix q_{ij} with zeros only on the diagonal. Then $Y(I) = \hat{p}$ for every I with two or more nodes, so that only the unit sets i must be checked. Conditions (8) now reduce to the requirement $b_i + d_i < \hat{p}$ for every node i. An important application involves choosing all b_i and d_j equal to one, so that $M = \hat{p} = N$ and $\mathbf{q}$ must be reduced to a doubly stochastic matrix $\mathbf{p}^*$. But (7) says that $X(I)$ and $Y(I)$ are just the sizes $|I|$ and $|J(I)|$ of I and $J(I)$, respectively. A result from matching theory, the "marriage theorem,"[9] now shows that the conditions $|I| \leqslant |J(I)|$ from (7) are equivalent to $\mathbf{q}$ having some positive principal diagonal, and that $|I| < |J(I)|$ from (8) imply that each positive element of $\mathbf{q}$ lies on a positive principal diagonal. These are the conditions cited by Sinkhorn and Knopp[5] for the doubly stochastic case. In a typical case, the modified labeling method will ordinarily be easier to apply than (6)–(8) in testing for existence of a solution.

The growth factors $\mathbf{E}$ and $\mathbf{F}$ are essentially unique, except for the possibility of scaling them along rays as described earlier. Indeed, let $\hat{\mathbf{E}}$ and $\hat{\mathbf{F}}$ satisfy $p_{ij} = q_{ij}E_iF_j = q_{ij}\hat{E}_i\hat{F}_j$ for all i and j. Then $\hat{E}_i/E_i = F_j/\hat{F}_j = a$ whenever q_{ij} is not zero, and the factor a is independent of i and j since the network is connected. When the problem decomposes into independent Kruithof subproblems, the network breaks into separate connected components, each with a separate scaling factor a. The case that $\mathbf{p}$ and $\mathbf{q}$ are symmetric implies that $\mathbf{E} = a\mathbf{F}$, so that the growth factors may be scaled to satisfy $\mathbf{E} = \mathbf{F}$ uniquely.

IV. CONVERGENCE CONSIDERATIONS

In iterating (4) and (5) to solve for $\mathbf{p}^*$, a good convergence test can be based on the norm consisting of the sum of the absolute values of the differences between the left and right sides of (1) and (2):

$$g = \sum_i |\, b_i - E_i \sum_j q_{ij}F_j \,| + \sum_j |\, d_j - F_j \sum_i q_{ij}E_i \,|. \tag{11}$$

Clearly, g is the net error in traffic units and goes to zero as $\mathbf{p}$ goes to the solution $\mathbf{p}^*$; we can show that, in fact, g does so monotonically. Indeed, g is continuous, convex, and piecewise linear in $\mathbf{E}$, with E_i-derivatives of the form:

$$-\sum_j q_{ij}F_j \left[\operatorname{sgn}\left(b_i - E_i \sum_{j'} q_{ij'}F_{j'} \right) + \operatorname{sgn}\left(d_j - F_j \sum_{i'} q_{i'j}E_{i'} \right) \right],$$

and this expression always takes the sign of $E_i \sum q_{ij}F_j - b_i$ or else vanishes. Thus g can only decrease or remain constant as E_i is increased or decreased to satisfy (4) and g achieves its minimum, for any fixed value of $\mathbf{F}$, when each E_i is given by (4). A similar discussion holds with respect to the F_j-derivative of g. Accordingly, when g becomes small during iteration, it will remain so, and it is appropriate to stop.

During the process of iterating to compute $\mathbf{E}$ and $\mathbf{F}$, we can accumulate a value of g with very little additional effort. Specifically, when all the F_j have a new value, the second term of (11) vanishes. Proceeding to update the E_i, we compute $\sum q_{ij}F_j$ for use in (4). With two more additions and one multiplication, we get the corresponding contribution to g in the first term of (11). By the time all new E_i are computed, a value of g for $\mathbf{F}$ and the old $\mathbf{E}$ is available. Clearly, the iteration may be interpreted as a relaxation scheme to minimize g by cyclically minimizing over the E_i and F_j.

Since nonexistence of a solution $\mathbf{p}^*$ when (6) holds can only accompany zero elements in $\mathbf{q}$, one might attempt to force a solution by substituting a small positive value for each zero. In general, this is a terrible idea; Sinkhorn,[4] for instance, shows some examples of pathological behavior associated with such schemes. The iteration process will seek to get significant flow on some edges with small q_{ij} by using very large growth factors. The solution $\mathbf{p}^*$ will not greatly resemble $\mathbf{q}$ and the iteration will ordinarily converge very slowly. Indeed, slow convergence is a possible warning that the problem is ill-conditioned in some way. For such cases, it is prudent to check the "existence margin" by setting successively larger elements of $\mathbf{q}$ to zero and running the existence test until it fails—just the reverse of adding small positive terms. The labeling method is sufficiently economical of computing time that it may be repeated often in preference to performing a great many iterations.

In general, ill-conditioning occurs when some of the total traffic values b_i and d_j are not particularly consistent with one another. That is, when some collection of small elements q_{ij} are set to zero, (1)–(2) no longer have any solution for which p_{ij} vanishes if and only if q_{ij} does. The limit as these elements approach zero may not exist, or it may depend on the specific way in which they vanish. In short, $\mathbf{p}^*$ is not necessarily continuous in the elements of $\mathbf{q}$ at the boundaries of the feasible region. On the other hand, $\mathbf{p}^*$ is provably continuous in the row and column sums, or any other constraint levels, so that adjusting them to achieve consistency at the boundary is a stable procedure. Thus, the proper way to treat ill-conditioning is by readjusting some values of b_i and d_j to make them more consistent, though it may not be obvious how to do this. We see later that there is an easy way to extend the Kruithof method so that it automatically allocates traffic among the rows or columns of prespecified aggregations in a consistent and reasonable way.

It is also possible for a quite reasonable problem to converge at a very slow rate. For example, a problem may decompose into multiple independent subproblems, whose network components are only connected by way of s and t. Now, each of the subproblems may be well-behaved, so that the overall iteration process converges rapidly. Nevertheless, when a few small positive values of q_{ij} are introduced to couple the subproblems, convergence will be rapid at first and then become rather slow, as a rule. Moreover, the solution to which the problem converges is a sensible one that differs only slightly from the decoupled case. This model could apply to two or more countries, for example, with much more traffic internally than across their borders.

What causes the above difficulty is the difference in degree of uniqueness between coupled and decoupled cases. For n subproblems, n independent arbitrary scaling factors a will appear in the general solution. The actual values they assume will be determined by the initial values assigned to $\mathbf{E}$ or $\mathbf{F}$. This effect appears as some arbitrariness in the relative sizes of the E_i for those portions of $\mathbf{E}$ associated with the various subproblems. For the coupled case, only a single overall scaling factor is appropriate; the portions of $\mathbf{E}$ from the different subproblems must now be scaled in a correct ratio to each other. In the rapidly convergent phase, each subproblem is solved to within its scale factor; the slow phase corresponds to a process of adjusting scale factors to account for small coupling terms.

The slowly convergent phase may be shortened appreciably by means of standard numerical schemes for acceleration of convergence. Wynn's algorithm,[10] in particular, has been employed successfully to project the values of the E_i for successive steps, in order to estimate their limit. The strategy is to calculate the norm g for the projected $\mathbf{E}$ and, when this is much less than the current error g (one-fiftieth, say), restart the iteration from the projected value. Additional computational effort and program steps for a convergence acceleration option in Kruithof's method are minor. Of course, acceleration techniques cannot rescue a truly ill-conditioned case, where the existence margin is small. All this is not meant to imply that slow convergence is the rule with the Kruithof method. In fact, some rather large examples, involving several hundred rows and columns, have been solved quite readily.

V. VARIOUS EXTENSIONS

An immediate generalization of the basic Kruithof scheme would be to stratify the traffic data in more than two dimensions. Besides the originating and terminating points i and j, other indices k, l, m $\cdots$ might specify time of day, week, or year, type of traffic (business or residential, for instance), and so on. A three-dimensional case of the general Kruithof problem then might take the specific form:

$$p_{ijk} = q_{ijk} E_i F_j G_k \tag{12}$$

$$b_i = \sum_{j,k} p_{ijk} = E_i \sum_{j,k} q_{ijk} F_j G_k \tag{13}$$

$$d_j = \sum_{i,k} p_{ijk} = F_j \sum_{i,k} q_{ijk} E_i G_k \tag{14}$$

$$f_k = \sum_{i,j} p_{ijk} = G_k \sum_{i,j} q_{ijk} E_i F_j. \tag{15}$$

Reduction to the standard case in the appendix proceeds as before. We define events $e = (i, j, k)$ and constraints c corresponding to each value of i, j, and k, while $\mathbf{p}$ and $\mathbf{q}$ are normalized to probabilities by dividing by the sums of their elements.

The proofs of existence, uniqueness, and convergence in the appendix still hold for this case. In particular, a solution $\mathbf{p}^*$ of (12)–(15) exists if and only if (13)–(15) have a solution p_{ijk} that is positive or zero accordingly as q_{ijk} is positive or zero. The norm g consisting of the sum of absolute values of differences between left and right sides in (13)–(15) is still net error in traffic units, though it no longer decreases monotonically with each new value of E_i, F_j, or G_k. Nevertheless, g

goes to zero as **p** goes to **p*** and it is still a reasonable indicator of convergence. When **q** is strictly positive, the interior solution $p_{ijk} = b_i d_j f_k / \hat{p}^2$ of (13)–(15) demonstrates that **p*** also exists, provided that the b_i, d_j, and f_k each add up to $\hat{p}$. A more general existence test would involve solving the linear programming problem described in the appendix, with appropriate precautions against being misled by the effects of any roundoff error. Network flow models no longer apply, and integer solutions need not occur in the case of integer constraints.

Another three-dimensional example of the general problem might be as follows:

$$p_{ijk} = q_{ijk} E_{ij} F_{jk} G_{ik} \tag{16}$$

$$b_{ij} = \sum_k p_{ijk} = E_{ij} \sum_k q_{ijk} F_{jk} G_{ik} \tag{17}$$

$$d_{jk} = \sum_i p_{ijk} = F_{jk} \sum_i q_{ijk} E_{ij} G_{ik} \tag{18}$$

$$f_{ik} = \sum_j p_{ijk} = G_{ik} \sum_j q_{ijk} E_{ij} F_{jk}, \tag{19}$$

and the same sort of discussion applies to this case as to (12)–(15). For four or more dimensions, the reader should have no difficulty creating a great many extensions of this general class, such as $p_{ijkl} = q_{ijkl} B_{ij} C_{jk} D_{kl} E_{il} F_{ik} G_{jl}$. The rule is to multiply $q_{ijkl\ldots}$ by one factor $E_{ij\ldots}$ for each constraint in which $p_{ijkl\ldots}$ appears, and then solve each constraint for its factor by dividing into the constraint level $b_{ij\ldots}$. In higher dimensions, the number of elements in **p** and **q** grows much faster than the numbers of constraints and multipliers. The numbers of the latter thus remain reasonable, if only to stay within storage and computational limits on the former. The corresponding linear program to test for existence will therefore have a basis of reasonable size, as well.

Another class of extensions involves specifying less about total traffic quantities in the two-dimensional case (or any other dimension, using the previous generalization). That is, various sets I or J of originating or terminating points i or j may be lumped together, with only their total traffic $\tilde{\mathbf{b}}$ and $\tilde{\mathbf{d}}$ to be supplied externally. Now (1)–(2) yield the following constraints:

$$\tilde{b}_I \equiv \sum_I b_i = \sum_I \sum_j p_{ij} \tag{20}$$

$$\tilde{d}_J \equiv \sum_J d_j = \sum_J \sum_i p_{ij}. \tag{21}$$

Results on the general Kruithof problem in the appendix show that $E_i = \tilde{E}_I$ and $F_j = \tilde{F}_J$ in this case, so that (3) becomes:

$$p_{ij} = q_{ij} E_i F_j = q_{ij} \tilde{E}_I \tilde{F}_J, \tag{22}$$

assuming each i belongs to one I and each j to one J.

To solve for $\bar{\mathbf{E}}$ and $\bar{\mathbf{F}}$, we can collapse the problem to a simpler form: add together all rows i in I and all columns j in J to work with reduced matrices $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$, as follows:

$$\tilde{p}_{IJ} \equiv \sum_I \sum_J p_{ij} \qquad \tilde{q}_{IJ} \equiv \sum_I \sum_J q_{ij}. \tag{23}$$

Now (20)–(22) are reduced to exactly the same form as (1)–(3):

$$\tilde{b}_I = \sum_J \tilde{p}_{IJ} \qquad \tilde{d}_J = \sum_I \tilde{p}_{IJ} \qquad \tilde{p}_{IJ} = \tilde{q}_{IJ} \bar{E}_I \bar{F}_J, \tag{24}$$

but with many fewer constraints to be satisfied. Note that this procedure of aggregating traffic nodes may absorb some elements of $\mathbf{q}$ that were zero or small, in order to ameliorate ill-conditioning. Indeed, the corresponding disaggregation formula (22), to be used after we have solved (24), simply allocates traffic p_{ij} to the element q_{ij} in proportion to its relative contribution to $\tilde{q}_{IJ}$ in (23). This in turn automatically shares out $\tilde{b}_I$ and $\tilde{d}_J$ among their b_i and d_j in a consistent manner, as mentioned earlier. As another of its virtues, aggregation is a smoothing process that can cut down the effects of errors in the predictions of total traffic by reducing the number of independent parameters. The reduction in manual and computational effort is also an evident advantage. To organize the computation in an efficient manner, we would start with two tables, $I(i)$ and $J(j)$, that assign each point i or j to its appropriate aggregate. Then we run through the pairs (i, j), adding each q_{ij} into its correct $\tilde{q}_{I(i)J(j)}$. After we solve (24) for $\bar{\mathbf{E}}$ and $\bar{\mathbf{F}}$, the answer is just $p_{ij} = q_{ij} \bar{E}_{I(i)} \bar{F}_{J(j)}$ from (22). Existence testing can be performed directly on $\tilde{\mathbf{q}}$, $\tilde{\mathbf{b}}$, and $\tilde{\mathbf{d}}$ with the labeling method.

The scheme above illustrates a sense in which a "separability" property can be introduced, similar to what Kruithof sought. There is no real need to require that different I or J be disjoint. If overlap is permitted, then q_{ij} would be multiplied by $\bar{E}_I$ for each I that contains i, and by $\bar{F}_J$ for each J containing j. Collections of rows or columns can now be aggregated only if they all lie in the same sets I or J. Since some p_{ij} may now be counted twice, the constraints can no longer be interpreted as conservation laws for a flow, and the existence test becomes a linear program, as described in the appendix.

Another way of specifying less in the Kruithof problem is to leave some rows or columns unconstrained. At such i and j, we can specify the growth factors E_i and F_j arbitrarily; Bear[2] has considered choosing these multipliers to be one. One advantage of not constraining is that any element q_{ij} for which row i and column j are not constrained plays no part in the solution process and may be set to zero for convenience. For computational efficiency, we multiply each of these rows by its fixed growth factor and add together all such rows to form a single new row; similarly, all unconstrained columns combine. Of course, fewer specifications may be introduced in higher-dimensional schemes as well.

Anyone can tailor their own ad hoc constraints to account for additional knowledge of the future. For instance, it may be known that certain items of point-to-point traffic p_{ij} are growing considerably faster than the other, more typical items in their rows and columns. Then a constraint may be created to fix the projected value of the sum of all such items. This produces one more growth factor to be multiplied into q_{ij} for these selected items only. Similar treatment may be given to a class of items that have slower than normal growth, that decrease, or that just vanish. The general Kruithof problem defined in the appendix includes all such cases, as well as any other linear constraints that may need to be introduced.

VI. SUMMARY AND COMMENTS

In this study, we have sought to give a unified view of Kruithof's teletraffic projection method, including theoretical aspects as well as practical details for its implementation. The mathematics in the appendix treats the theory of a general Kruithof problem. Necessary and sufficient conditions for existence and convergence of its solution are derived, along with proofs of uniqueness and continuity. The main text considers special cases of this problem that are of particular interest. Schemes for existence and convergence testing and for handling slow convergence are discussed. We conclude by trying to place this projection scheme in the context of the Bell System planning function.

An early and important step in the Bell System planning process is that of predicting future demand for the various services offered. By their nature, such projections can be quite uncertain, since they will include cumulative effects of several years' fluctuations in the United States and world economy, for instance. Indeed, analysis of time series of typical traffic data[11] indicates that about five percent per year of random error remains in even the best projections, and must be regarded as inherently unpredictable. Nevertheless, a strategy is available to cope with such uncertainties, for the purposes of planning.

The fundamental assumption required in this strategy is that traffic increases monotonically with time. First a plan can be generated, based upon some "best guess" of the demand profile over time. From year to year, the time scale of the plan can then be corrected to match up the originally projected demand with actual values or better estimates, based upon more recent data. In effect, a parameter such as total traffic is thus used as a new independent variable in the plan, while time is a dependent variable that absorbs much of the economic fluctuations and other error. However, this leads us to view the overall process of planning as a system, rather than a collection of independent modules, one of which is projection. We see that a "sliding time scale" approach to planning now places a premium not so much on absolute accuracy of traffic predictions over time as on "relative accuracy" or consistency and uniformity among the various traffic quantities that are projected.

A typical plan may be based upon many thousands or tens of thousands of projected traffic items. We wish to predict these quantities such that a single readjustment of the plan time scale (based on some total traffic measure, for instance) can do a reasonable job of correcting for the error in each item. This requires that projection be done in such a way that the prediction errors in individual traffic items will tend to be highly correlated. Thus, a scheme which projected individual time series for each separate traffic quantity, for example, might give the best absolute accuracy for each item, but still be unsuitable for planning purposes because the noise components in these time series would tend to be independent. At the opposite extreme, initial measurements of all traffic quantities might simply be increased by a single overall growth factor for each year under study. This would produce very strong correlations but fails to take into account detailed knowledge of growth patterns, say, for separate portions of a study area.

The general Kruithof method offers many middle roads. Any collection of average or overall traffic quantities $\mathbf{b}$ may be predicted externally (from time series, for example, or market surveys). As shown in the appendix, the remaining items can then be projected to be consistent with whatever is given. Inserting a great many external predictions introduces more detailed knowledge, but reduces the correlation. Supplying fewer external specifications yields stronger correlations, at some loss in accuracy; various tradeoffs are possible.

All the schemes of Kruithof type act to minimize the net information change in the projection, subject to those external constraints being enforced. This gives them the remarkable properties called "reversibility" and "divisibility" by Kruithof. Essentially, all that is lost of the original data $\mathbf{q}$ in projecting it to future values $\mathbf{p}$ is whatever is inherent in the externally provided average quantities $\mathbf{b}$. Supplying new values $\mathbf{b}'$ for these quantities will thus allow us to continue the projection from $\mathbf{p}$ to another year or recover the base data $\mathbf{q}$ exactly. Effectively, Kruithof's method is able to resolve $\mathbf{q}$ and $\mathbf{p}$ into a part which determines some arbitrarily chosen system of average quantities $\mathbf{b}$ that are to be changed and an "orthogonal" part that does not change. This latter part is essentially the equivalence class $C(\mathbf{Q})$ discussed in the appendix.

Kruithof's original proposal, and much of the subsequent work on the subject, is concerned with projecting two-dimensional arrays p_{ij} of traffic data. In this case, the most natural overall quantities to be specified externally are the sums of rows i and columns j in $\mathbf{p}$. For example, supplying two hundred parameters for a 100×100 matrix would suffice to project a total of ten thousand items. An option is to aggregate rows and columns, perhaps in collections of average size four, so that only fifty external sums are needed. This is a tradeoff that increases correlations but may decrease accuracy. Meanwhile, it can alleviate possible ill-conditioning, reduces manual and computational effort, and still projects ten thousand items. Examples of data organized in three or more dimensions are also known in Bell System planning. Kruithof's method generalizes to such cases without any particular difficulty.

VII. ACKNOWLEDGMENTS

Kruithof's method was first brought to my attention by Pierre L. Bastien; many of the ideas here originated in discussions with him. Various users of Kruithof's method in Bell System applications have pointed out interesting properties and extensions. Several colleagues were kind enough to read and comment upon draft copies; P. Kazakos was particularly helpful in supplying references.

APPENDIX

In this appendix, we define the general case of the Kruithof problem and derive various properties. In particular, necessary and sufficient conditions for existence and convergence of solutions are developed, and uniqueness and continuity are proved. To begin with, consider two probability distributions, P_e and Q_e, over the same finite set of disjoint events $\{e\}$, so that:

$$\sum_e P_e = 1 \qquad P_e \geqslant 0 \tag{25}$$

$$\sum_e Q_e = 1 \qquad Q_e \geqslant 0. \tag{26}$$

The information content of a probability, in appropriate units, is minus its logarithm. Thus the change in information from Q_e to P_e is just $\log P_e - \log Q_e = \log(P_e/Q_e)$. The average of this information change is defined as:

$$K(\mathbf{P}, \mathbf{Q}) \equiv \sum_e P_e \log(P_e/Q_e), \tag{27}$$

which will be interpreted as a measure of how close distribution $\mathbf{P}$ is to distribution $\mathbf{Q}$. A simple example is the case that all probabilities Q_e are equal; now K reduces to a linear function of the entropy of distribution $\mathbf{P}$. Thus entropy measures departure from the equiprobable case (corresponding to classical equilibrium). The expression (27) goes by several names in the literature; for instance, Kullback distance, I-divergence, relative entropy, discrimination information, and Gibbs free energy.

As a distance measure, K has two desirable properties: it is not negative and vanishes only when $\mathbf{P} = \mathbf{Q}$. Two other properties, symmetry and the triangle inequality, are lacking; Csiszar[7] points out, however, that laws analogous to the parallelogram identity and Pythagoras' theorem do hold, with K playing the role of squared length. To show that K is nonnegative, first note that $F(X) \equiv \log(1/X)$ is strictly convex, since $F'' = 1/X^2 > 0$. Defining $X_e \equiv Q_e/P_e$ and using convexity in (27) yields the inequality:

$$K = \sum_e P_e F(X_e) \geqslant F\left(\sum_e P_e X_e\right) = F(1) = 0, \tag{28}$$

with equality only if all $P_e F(X_e)$ vanish, in which case $\mathbf{P} = \mathbf{Q}$ from (25)–(26).

When Q_e is positive and P_e approaches zero, $P_e \log(P_e/Q_e)$ goes to a zero limit; to ensure continuity, we will define it to vanish at $P_e = 0$, even for the case $Q_e = 0$. If some $Q_e = 0$ for positive P_e, then K is infinite. Confining our attention to $\mathbf{P}$ that are not infinitely far from $\mathbf{Q}$, P_e must vanish whenever Q_e does, so that e is an event of probability zero. With no loss of generality, such e may be excluded from the set $\{e\}$ of events for now, so that all Q_e are positive.

Consider the problem of minimizing $K(\mathbf{P}, \mathbf{Q})$ for fixed $\mathbf{Q}$, over all distributions $\mathbf{P}$ subject to (25) and a finite set $\{c\}$ of arbitrary linear constraints having the general form:

$$\sum_e P_e A_{ec} = B_c. \tag{29}$$

This amounts to finding the distribution $\mathbf{P}$ that is closest to $\mathbf{Q}$ on the intersection (denoted $S(\mathbf{B})$, or just S) of the positive orthant $\mathbf{P} \geqslant 0$ and the hyperplanes (29), which prescribe that certain averages over $\mathbf{P}$ take on the values B_c. (To simplify the notation, assume that the equality in (25) is designated $\hat{c}$ and is included among the constraints c.) Observe that S is a compact convex polytope, so that the continuous function K achieves its minimum value on S, whenever S is not empty. Further, this minimum occurs at a unique point $\mathbf{P}^*$ in S, and there are no other local minima of K, since it is strictly convex. To see this, note that the matrix of second partial derivatives of K with respect to $\mathbf{P}$ is diagonal and positive definite. We assume from now on that S contains more than one point.

Suppose that S contains an interior point $\mathbf{P}^{Int}$ (that is, no P_e^{Int} vanishes); then $\mathbf{P}^*$ is also an interior point. Indeed, consider any boundary point $\mathbf{P}^{Bdy}$ and the line $\mathbf{P}^{Bdy}\mathbf{P}^{Int}$ joining it to $\mathbf{P}^{Int}$. The gradient of K has components $1 + \log(P_e/Q_e)$ that become arbitrarily large negative on some neighborhood of $\mathbf{P}^{Bdy}$ for those P_e^{Bdy} that vanish, while all other components remain bounded. Thus, a segment of $\mathbf{P}^{Bdy}\mathbf{P}^{Int}$ containing $\mathbf{P}^{Bdy}$ can be found along which K decreases toward the interior, and $\mathbf{P}^{Bdy}$ cannot be a minimum point of K. Since the gradient of K is continuous on the interior of S, an interior $\mathbf{P}^*$ is a stationary point of K and, by strict convexity, there are no other stationary points of K on S.

When $\mathbf{P}^*$ is an interior point, the minimization problem can be solved by using stationarity. Specifically, adjoin constraints (29) to K with multipliers v_c to form the Lagrangian function $L(\mathbf{P}, \mathbf{v})$, as follows:

$$L \equiv \sum_e P_e \left[\log(P_e/Q_e) - \sum_c A_{ec} v_c \right] + \sum_c B_c v_c. \tag{30}$$

Now L is strictly convex on the positive orthant $\mathbf{P} \geqslant 0$ and becomes infinite if any P_e does, while its gradient is negative infinite on the boundaries. Thus it achieves a unique minimum over $\mathbf{P} \geqslant 0$ at some interior stationary point $\hat{\mathbf{P}}(\mathbf{v}) > 0$ for any fixed values of v_c. This point is found by requiring the P_e-derivative of L to vanish for each e, yielding the following necessary conditions:

$$w_e \equiv \log(\hat{P}_e/Q_e) = \sum_c A_{ec} v_c - 1. \tag{31}$$

Setting $V_c \equiv \exp(v_c)$ in (31), except for the constraint $\hat{c}$ corresponding to (25) with $V_{\hat{c}} \equiv \exp(v_{\hat{c}} - 1)$, now yields

$$\hat{P}_e = Q_e \exp(w_e) = Q_e \prod_c V_c^{A_{ec}}, \tag{32}$$

with all V_c strictly positive, so that $\hat{\mathbf{P}}(\mathbf{v})$ cannot be on the boundary.

Suppose some v_c are found such that $\hat{\mathbf{P}}(\mathbf{v})$ satisfies the constraints (29) and thus lies in S. Then since $L = K$ on S, $\hat{\mathbf{P}}(\mathbf{v})$ is a stationary point of K on S, and hence is the unique minimum point $\mathbf{P}^*$. Conversely, the linear program of minimizing dK for all small variations $d\mathbf{P}$ that satisfy (29) has (31) as its dual constraints. Since the primal problem has $d\mathbf{P} = 0$ as an optimum at $\mathbf{P}^*$, the dual is feasible there, sc that $\mathbf{v}$ can be found to satisfy (31) at $\mathbf{P}^*$. The point of all this reasoning is that a solution to (29) can be found with the specific product form (32) if and only if S possesses an interior. When such a solution exists, it is also the unique minimum point $\mathbf{P}^*$ of K over S. We can now define the Kruithof problem, in general, as that of finding the factors V_c in (32) so as to satisfy the constraints (29).

Kruithof's "reversibility" property amounts to symmetry of the "closeness" relation: whenever $\mathbf{P}$ is closest to $\mathbf{Q}$, then $\mathbf{Q}$ is closest to $\mathbf{P}$ in the same sense. That is, we define some new constraint levels:

$$\hat{B}_c \equiv \sum_e Q_e A_{ec} \tag{33}$$

and seek a distribution $\mathbf{R}$ to minimize:

$$\hat{K} \equiv K(\mathbf{R}, \mathbf{P}) = \sum_e R_e \log(R_e/P_e) \tag{34}$$

over the set $S(\hat{\mathbf{B}})$ defined by linear constraints:

$$\sum_e R_e A_{ec} = \hat{B}_c \qquad R_e \geqslant 0. \tag{35}$$

Defining $\hat{V}_c \equiv 1/V_c$, we can write Q_e from (32) in the product form:

$$Q_e = P_e \prod_c \hat{V}_c^{A_{ec}}. \tag{36}$$

From (26), (33), and (36), $\mathbf{Q}$ is an interior point of $S(\hat{\mathbf{B}})$ having the product form (32), so that $\mathbf{R} = \mathbf{Q}$ is the unique minimum solution for $\mathbf{R}$, and $\mathbf{Q}$ is closest to $\mathbf{P}$. Kruithof's "divisibility" property is just transitivity of the closeness relation: when $\mathbf{P}$ is closest to $\mathbf{Q}$ and $\mathbf{R}$ is closest to $\mathbf{P}$, then $\mathbf{R}$ is closest to $\mathbf{Q}$. Specifically, we choose an arbitrary constraint vector $\hat{\mathbf{B}}$ in (35), such that $S(\hat{\mathbf{B}})$ has an interior point, and seek $\mathbf{R}$ to minimize $\hat{K}$ in (34) over $S(\hat{\mathbf{B}})$. The solution for $\mathbf{R}$ is the right side of (36) for some $\hat{\mathbf{V}}$, and substituting for $\mathbf{P}$ from (32) yields:

$$R_e = Q_e \prod_c (\hat{V}_c V_c)^{A_{ec}}, \tag{37}$$

which again has the product form (32). Thus $\mathbf{R}$ minimizes $K(\mathbf{R}, \mathbf{Q})$ over $S(\hat{\mathbf{B}})$ and is closest to $\mathbf{Q}$ in this sense because it is closest to $\mathbf{P}$.

Symmetry and transitivity show that "closest to" is an equivalence relation determined by the particular matrix $[A_{ec}]$. This relation partitions the set of positive distributions over $\{e\}$ into equivalence classes. Each class $C(\mathbf{Q})$ can be generated from any one of its members $\mathbf{Q}$ by using (32): choose all positive values of the V_c for $c \neq \hat{c}$ and scale the resulting values $\hat{\mathbf{P}}$ as necessary to meet the normalization condition (25). Since the column of $[A_{ec}]$ corresponding to $\hat{c}$ is all ones, $V_{\hat{c}}$ appears linearly in (32), and the scaling above represents a particular choice of that variable. Uniqueness says that each constraint vector $\mathbf{B}$ is achieved at most once in each class, while the existence condition asserts that $\mathbf{B}$ is achieved in every class if it is achieved in one class. A natural mapping $\mathbf{B} = f(\mathbf{P})$, namely the linear mapping (29), takes any $C(\mathbf{Q})$ into the set E of constraint vectors $\mathbf{B}$ for which $S(\mathbf{B})$ has an interior. Clearly, f is continuous and is one-to-one and onto E from the existence and uniqueness results. We will see later that f is one-to-one on the closure of $C(\mathbf{Q})$, which is compact from (25). (However the closure is not necessarily an equivalence class.) It follows that f is a homeomorphism of $C(\mathbf{Q})$ and E. In particular, $\mathbf{P}^*$ is a uniformly continuous function of the constraint vector $\mathbf{B}$ on E and its closure.

A useful result follows from the linear relation between $\mathbf{v}$ and $\mathbf{w}$ in (31). Specifically, let $\mathbf{R}$ be any solution of (25) and (29), multiply R_e by w_e, and sum over e, using (29) and (31) to obtain

$$\sum_e R_e w_e = \sum_e R_e \left[\sum_c A_{ec} v_c - 1 \right] = \sum_c B_c v_c - 1 \tag{38}$$

by exchanging the order of summation. It is easy to show from (31) that the left side of (38) is just $K(\mathbf{R}, \mathbf{Q}) - K(\mathbf{R}, \hat{\mathbf{P}})$, while the right side does not depend on the particular choice of $\mathbf{R}$ in $S(\mathbf{B})$, so that

$$K(\mathbf{R}, \mathbf{Q}) + K(\mathbf{P}, \hat{\mathbf{P}}) = K(\mathbf{R}, \hat{\mathbf{P}}) + K(\mathbf{P}, \mathbf{Q})$$

for all $\mathbf{R}$, $\mathbf{P}$ in S and all $\hat{\mathbf{P}}$, $\mathbf{Q}$ in C. But now the choice $\mathbf{P} = \mathbf{P}^* = \hat{\mathbf{P}}$ yields an example of the Pythagorean theorem noted by Csiszar:

$$K(\mathbf{R}, \mathbf{Q}) = K(\mathbf{R}, \mathbf{P}^*) + K(\mathbf{P}^*, \mathbf{Q}). \tag{39}$$

Roughly, it says that the distance from $\mathbf{R}$ to $\mathbf{Q}$ breaks into a component within S from $\mathbf{R}$ to $C(\mathbf{Q})$ at $\mathbf{P}^*$ and an "orthogonal" component from $\mathbf{P}^*$ to $\mathbf{Q}$ within $C(\mathbf{Q})$.

Now we will investigate the nonlinear (Wolfe) dual problem to minimization of K on S. Let $H(\mathbf{v})$ be the minimum of $L(\mathbf{P}, \mathbf{v})$ over the positive orthant $\mathbf{P} \geqslant 0$, so that

$$H(\mathbf{v}) = L(\hat{\mathbf{P}}(\mathbf{v}), \mathbf{v}) \leqslant L(\mathbf{P}^*, \mathbf{v}) = K(\mathbf{P}^*, \mathbf{Q}) \leqslant K(\mathbf{R}, \mathbf{Q}), \tag{40}$$

which shows that $H(\mathbf{v})$ is bounded above if S is not empty. Substituting (31), (32), and (38) into (30) allows us to express H as a function of each of $\mathbf{v}$, $\mathbf{V}$, $\mathbf{w}$ and $\hat{\mathbf{P}}$ as follows:

$$\begin{aligned} H &= \sum_c B_c v_c - \sum_e Q_e \exp\left(\sum_c A_{ec} v_c - 1 \right) \\ &= 1 + \sum_c B_c \log(V_c) - \sum_e Q_e \prod_c V_c^{A_{ec}} \\ &= 1 + \sum_e [R_e w_e - Q_e \exp(w_e)] = 1 + \sum_e [R_e \log(\hat{P}_e/Q_e) - \hat{P}_e]. \end{aligned} \tag{41}$$

Direct differentiation of $H(\hat{\mathbf{P}})$ shows that it achieves a unique maximum over the positive orthant at $\hat{\mathbf{P}} = \mathbf{R}$ where $H = K(\mathbf{R}, \mathbf{Q})$. (Break H into a linear part for those components of $\mathbf{R}$ that vanish and a strictly concave part.) Indeed, if $H(\hat{\mathbf{P}})$ goes to $K(\mathbf{R}, \mathbf{Q})$ on $\hat{\mathbf{P}} \geqslant 0$, we can conclude that $\hat{\mathbf{P}}$ approaches $\mathbf{R}$. Now the difference between $K(\mathbf{R}, \mathbf{Q})$ and $H(\hat{\mathbf{P}})$ achieves a unique minimum of zero at $\hat{\mathbf{P}} = \mathbf{R}$:

$$K - H = \sum_e [R_e \log(R_e/\hat{P}_e) + \hat{P}_e] - 1 = K(\mathbf{R}, \hat{\mathbf{P}}) + \sum_e \hat{P}_e - 1. \tag{42}$$

Whenever this expression vanishes for some $\hat{\mathbf{P}}$ of the form (32), we have $K(\mathbf{P}^*, \mathbf{Q}) = K(\mathbf{R}, \mathbf{Q})$ from (40), and thus $\mathbf{P}^* = \mathbf{R}$ by uniqueness of the minimum.

Suppose that values of $\hat{\mathbf{P}}(\mathbf{v})$ of the form (32) approach some limit $\mathbf{R}$ on the closure of $C(\mathbf{Q})$. Such $\mathbf{R}$ satisfies (25) and thus (29) for the constraint vector $\mathbf{B} = f(\mathbf{R})$. Now $K - H$ becomes arbitrarily small, and we conclude that $\mathbf{R}$ is the $\mathbf{P}^*$ that minimizes K. This shows that the closure of $C(\mathbf{Q})$ consists of points $\mathbf{P}^*$ that minimize $K(\mathbf{P}, \mathbf{Q})$. The uniqueness of the minimum then says that the mapping f is one-to-one on the closure, as promised earlier. Choosing $\mathbf{R} = \mathbf{P}^*$ in (41) shows that $H(\mathbf{P}^*) = K(\mathbf{P}^*, \mathbf{Q})$ even when $H(\hat{\mathbf{P}})$ is extended to the closure of $C(\mathbf{Q})$ by continuity. Finally, the result (39) again follows by setting $\hat{\mathbf{P}} = \mathbf{P}^*$ in (42).

The nonlinear dual problem consists of maximizing H over all $\mathbf{v}$, and thus over positive $\mathbf{V}$, or over the linear affine space of values $\mathbf{w}$ generated by (31). By its construction, H is concave in $\mathbf{v}$, while differentiation shows that it is strictly concave in each individual v_c. Setting the V_c-derivatives of H to zero yields necessary conditions for a stationary maximum:

$$B_c = \sum_e Q_e A_{ec} \prod_{c'} V_{c'}^{A_{ec'}}. \tag{43}$$

These are the same relations that would be obtained by substituting (32) into (29), namely, the general Kruithof problem. One possible scheme for solving eqs. (43) would be by relaxation: choose some variable V_c and adjust it to maximize H with all other variables $V_{c'}$ held fixed; then choose some other variable and repeat, cycling through all V_c infinitely often. (When H is bounded above by $K(\mathbf{R}, \mathbf{Q})$ in (40), a value of V_c to maximize H and satisfy (43) always exists uniquely, because $-H$ is strictly convex in each v_c and is arbitrarily large for large v_c.) If the values of $\hat{\mathbf{P}}(\mathbf{v})$ for the iterates $\mathbf{v}$ approach the limiting value $\mathbf{P}^*$, then the relaxation procedure represents a means of computing $\mathbf{P}^*$. More generally, any collection of constraints c could be solved simultaneously in (43), followed by another collection, and so on, so that each c appears infinitely often. Simultaneous solution of several constraints can be harder than the scheme of treating one variable at a time, however. Another possibility might be to solve (43) approximately for V_c, so that H increases at each step, but not necessarily to its exact maximum in V_c. The general relaxation procedure is just an attempt to maximize H over all variables by doing a few at a time. Such relaxation schemes are known[12] to converge to the maximum of a concave function under very general conditions.

In practice, A_{ec} will generally be a zero-one matrix, so that the powers of V_c in (32) do not become a nuisance. In such a case, the constraints (29) have a simple interpretation, since they assign probability B_c to the event c that is the disjoint union of those e for which $A_{ec} = 1$. Pursuing this view, the sum in (29) is taken over all events e included in c (denoted $e \subset c$) and the product in (32) is taken over all events c that include event e (denoted $c \supset e$). Substituting (32) into (29) now yields

$$B_c = \sum_{e \subset c} Q_e \prod_{c' \supset e} V_{c'} = V_c \sum_{e \subset c} Q_e \prod_{c \neq c' \supset e} V_{c'}, \tag{44}$$

as the form taken by (43) in this case. The relaxation iteration step now reduces to dividing B_c by the sum on the right in (44), in order to calculate the new value of V_c that maximizes H. Under certain weak restrictions on the sequence of variables chosen for iteration, it can be shown that H increases to the limit $K(\mathbf{P}^*, \mathbf{Q})$ and that $\hat{\mathbf{P}}(\mathbf{v})$ converges to $\mathbf{P}^*$. Specifically, assume that the iteration scheme includes an infinity of intervals of length M, for some sufficiently large M, in each of which H is maximized over every variable V_c at least once. Then the relaxation iteration converges if and only if S is nonempty; the limit is $\mathbf{P}^*$, which minimizes K over S and lies on the closure of $C(\mathbf{Q})$.

Indeed, with S nonempty, the consecutive values of H are nondecreasing but bounded above in (40). Thus H approaches some limit H^*, and the successive increases dH must eventually go to zero. Direct computation from (41) and (44) now yields the relation:

$$dH = B_c[dv_c - 1 + \exp(-dv_c)], \tag{45}$$

where dv_c is the corresponding change in v_c. Differentiation shows this expression to be strictly convex in dv_c (except in the trivial case $B_c = 0$), vanishing only at its minimum, namely at $dv_c = 0$. Thus each dv_c also goes to zero as H approaches H^*, so that $d\mathbf{w}$ goes to zero from (31). Consider the values of $\hat{\mathbf{P}}(\mathbf{v})$ that are obtained each time the constraint $\hat{c}$ corresponding to (25) is satisfied in one of the postulated intervals of length M. Since these values are confined to the simplex defined by (25), they have a subsequence that converges to some limit point $\mathbf{R}$. Now each member of the subsequence differs from a solution of constraint c by at most M changes, each of order $d\mathbf{w}$. It follows that the limit $\mathbf{R}$ will satisfy every constraint c. But then, from the discussion after (42), H goes to $K(\mathbf{R}, \mathbf{Q})$ on the subsequence and $\mathbf{R}$ is $\mathbf{P}^*$. Finally, H increases to its maximum over $\hat{\mathbf{P}} \geqslant 0$ for all iterates $\hat{\mathbf{P}}$, so that they converge to $\mathbf{P}^*$. Conversely, whenever the $\hat{\mathbf{P}}$ converge, the limit $\mathbf{R}$ satisfies all constraints, so that S is nonempty. Csiszar proves convergence for cylic iteration on collections of constraints, if each collection contains $\hat{c}$, by an elegant application of (39). In general, such cases do not include single-variable relaxation schemes, such as those treated above.

The artificial restriction that no Q_e may vanish can now be dropped, since (32) shows that P_e is zero whenever Q_e is in any case. The definition of S and its interior must be modified to account for all such conditions, of course. That is, (25) and (29) are supplemented by requirements that P_e vanish whenever Q_e does, while $P_e = 0$ makes $\mathbf{P}$ a boundary point of S only if Q_e is positive. Thus the Kruithof solution exists if and only if some $\mathbf{P}$ satisfies the constraints and vanishes for exactly the same events that $\mathbf{Q}$ does.

The question of whether the existence condition is met for a particular constraint vector $\mathbf{B}$ can be resolved, in principle, by constructing such an interior $\mathbf{P}$ with standard linear programming techniques.

Indeed, consider the problem of maximizing $h \geqslant 0$ subject to linear constraints:

$$\sum_{Q_e \neq 0} (T_e + h)A_{ec} + u_c = B_c \qquad T_e \geqslant 0 \qquad u_c \geqslant 0, \tag{46}$$

where each equality c has been written such that $B_c \geqslant 0$. The slack variables form the initial basis $u_c = B_c$, and a phase one procedure minimizes their sum. If not all u_c are forced to zero at optimum, then $S(\mathbf{B})$ is empty. Otherwise, a point in S has been constructed, so that the relaxation iteration will converge. In this case, all the u_c are dropped and phase two proceeds to maximize h. As soon as some step causes h to exceed zero, $P_e = T_e + h$ is the desired interior point. If the optimum still has $h = 0$, then $S(\mathbf{B})$ has no interior, and the iterative solution will not take the product form (32).

REFERENCES

1. J. Kruithof, "Telefoonverkeersrekening," De Ingenieur, *52*, No. 8(1937), pp. E15–E25.
2. D. Bear, *Principles of Telecommunication-Traffic Engineering*, London: Peter Peregrinus, 1976, Ch. 13.
3. Solomon Kullback, *Information Theory and Statistics*, New York: Wiley, 1959.
4. Richard Sinkhorn, "A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices," Ann. Math. Statist., *35*, No. 2(1964), pp. 876–879.
5. Richard Sinkhorn and Paul Knopp, "Concerning Nonnegative Matrices and Doubly Stochastic Matrices," Pacific J. of Math., *21*, No. 2(1967), pp. 343–348.
6. Henri Theil, *Statistical Decomposition Analysis*, Netherlands: North-Holland, 1972, Ch. 3.
7. I. Csiszar, "I-Divergence Geometry of Probability Distributions and Minimization Problems," Ann. Prob., *3*, No. 1(1975), pp. 146–158.
8. L. R. Ford, Jr. and D. R. Fulkerson, *Flows in Networks*, Princeton: Princeton University Press, 1962, Ch. 1.
9. Oystein Ore, *Graphs and Their Uses*, New York: Random House, 1963, Ch. 4.
10. Peter Wynn, "On a Device for Computing the $e_m(S_n)$ Transformation," Math. Comp., *10*, No. 54(1956), pp. 91–96.
11. William A. Gale, private communication and unpublished work.
12. S. Schechter, "Minimization of a Convex Function by Relaxation," Chapter 7 in *Integer and Nonlinear Programming*, Netherlands: North-Holland, 1970, J. Abadie, ed.